

## Essential Motions and Energetic Contributions of Individual Residues in a Peptide Bound to an SH3 Domain

Jiří Kolafa,\* John W. Perram,<sup>†</sup> and Robert P. Bywater<sup>‡</sup>

\*E. Hála Laboratory of Thermodynamics, Institute of Chemical Process Fundamentals, Academy of Sciences, CZ-16502 Praha, Czech Republic, <sup>†</sup>Mærsk McKinney Møller Institute for Production Technology, University of Southern Denmark—main campus Odense University, DK-5230 Odense M, and <sup>‡</sup>Biostructure Group, Medicinal Chemistry, Novo Nordisk A/S, DK-2760 Måløv, Denmark

**ABSTRACT** We have studied protein–ligand interactions by molecular dynamics simulations using software designed to exploit parallel computing architectures. The trajectories were analyzed to extract the essential motions and to estimate the individual contributions of fragments of the ligand to overall binding enthalpy. Two forms of the bound ligand are compared, one with the termini blocked by covalent derivatization, and one in the underivatized, zwitterionic form. The ends of the peptide tend to bind more loosely in the capped form. We can observe significant motions in the bound ligand and distinguish between motions of the peptide backbone and of the side chains. This could be useful in designing ligands, which fit optimally to the binding protein. We show that it is possible to determine the different contributions of each residue in a peptide to the enthalpy of binding. Proline is a major net contributor to binding enthalpy, in keeping with the known propensity for this family of proteins to bind proline-rich peptides.

### INTRODUCTION

Proteins are dynamic entities so that any attempt to understand their mechanism of action requires an analysis of their dynamic behavior while they are performing their allotted function, e.g., ligand binding. Having an accurate crystallographic or nuclear magnetic resonance (NMR) structure is a prerequisite for this, but it is certainly not enough. Although limited information about motions in the protein can be derived from certain types of crystallographic or NMR experiments, the picture that emerges from the use of these techniques is, to all intents and purposes, a static one. Molecular dynamics is a computational technique (McCammon et al. 1977; Brooks et al., 1983; Van Gunsteren and Berendsen 1987; Allen and Tildesley 1987) that enables the protein scientist to simulate the dynamic behavior of proteins and examine which motions are most critical for maintaining structure, unfolding, or refolding. Later applications of the technique have addressed issues such as the effect of mutations, docking to other proteins, or ligand binding (Peters et al., 1997; Peters and Bywater, 1999).

At the outset, one might consider what kind of dynamic behavior to expect from different types of protein having different functions. Some proteins like keratin in hair or silk fibroin do not bind a ligand at all but play a purely structural role. Other proteins have a more complicated function;

enzymes, for example, not only bind ligands but perform chemical reactions on them. This is usually accompanied by changes in the conformation of the protein and the ligand (Peters et al., 1997; Peters and Bywater, 1999). Allosteric enzymes, which respond to changes in conditions such as varying concentration of ligands or other effectors by changing their shapes and even function, will exhibit even more complicated motions. Somewhere between the functionally simpler structural proteins and the enzymes is a class of proteins that perform only a binding or transporter role for ligands. It would seem likely that the complexity of motions within these proteins would be intermediate also. The complexity of motions would most likely correlate with the degree of complexity of the function of the protein.

To quantify the degree of complexity and to be able to compare motions between, for example, cases where different ligands are bound to the same protein, or a mutation has been made, we developed a tool, described below, for analyzing the trajectories obtained from MD simulations. In this paper, we consider the motions in members of different families of proteins, starting with the peptide-binding protein SH3, which is present as a distinct domain, or folding unit, in many intracellular proteins and which mediates protein–protein interactions important in downstream signaling and cytoskeletal function (Kato et al., 1986; Potts et al., 1988; Nemeth et al., 1989; Seidel-Dugan et al., 1992; Koch et al., 1991; Musacchio et al., 1992; Pawson and Gish 1992; Pawson and Schlessinger 1993; Cicchetti et al., 1992; Weng et al., 1993; Barford et al., 1993; Gout et al., 1993; Liu et al., 1993; Taylor and Shalloway 1994; Fumagalli et al., 1994; Feng et al., 1994).

SH3 domains are small, water-soluble globular proteins typically constructed from about 60 amino acid residues consisting of 450 atoms. They have a compact, all- $\beta$  structure (SCOP fold type 21 [Murzin et al., 1995] and CATH

Received for publication 28 December 1999 and in final form 14 April 2000.

**Abbreviations used:** The standard single letter code for amino acid residue types is used throughout. MD, molecular dynamics; ED, essential dynamics; SH3, Src homology 3 domain; SH2, Src homology 2 domain.

Address reprint requests to Robert P. Bywater, Biostructure Group, Medicinal Chemistry, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Måløv, Denmark. Tel.: +45-4443-4530; Fax: +45-4443-4547; E-mail: byw@novo.dk.

© 2000 by the Biophysical Society

0006-3495/00/08/646/10 \$2.00

class 2.30.30.10. [Orengo et al., 1997]) approximately spherical in shape, but with a discernible ligand-binding region on the surface.

There is a very voluminous literature associated with SH3, and correspondingly, a large number of biophysical studies have been carried out, culminating in a number of crystal and NMR structures, some in the unliganded form and others in the ligand-bound form. Unfortunately, it is not always clear from these published studies whether the peptide ligand being studied is in a neutral ("capped," e.g., by acetylation at the N-terminus or amidation at the C-terminus) form or whether it is a zwitterion, as would seem to be the case. But the latter is not a good model for the interaction between SH3 and its target protein SH2, because the stretch of polypeptide in SH2 that is involved in binding will be flanked by a continuous peptide chain (i.e., the peptide is capped). Nevertheless, we have compared the dynamics and binding of both capped and zwitterion forms.

SH3 domains have previously been the subject of extensive MD simulations (Van Aalten et al., 1996) that were aimed at the development of protein-solvent simulation methodology and for studying protein dynamics using the technique of essential dynamics (Amadei et al., 1993). The conclusion was that a full solvent simulation was the recommended procedure for all simulations of water-soluble globular proteins, and, in this work, a cubic box of TIP3 waters was used. This, of course, requires higher overheads in the calculation than when using simpler approximations, but computational techniques embodied in the MACSIMUS (Macromolecule Simulation Software) package (Kolafa 1999), along with the use of a multiprocessor computer, reduce the computation time so that inclusion of a full solvent model is not a problem.

In earlier work, the SH3 protein 1shg (Noble et al., 1993), an unliganded form of the protein was studied (Van Aalten et al., 1996). The goal of the present studies is to compare and contrast the mode of binding and dynamic behavior of different peptide ligands bound to a similar protein, to study the essential motions in the ligand and ligand-binding site on the SH3.

One very useful objective of MD and ED studies on protein-ligand complexes is to be able to decide, based on the analysis of the dynamics, where in the ligand molecule it may be advantageous to "engineer in" or "engineer out" features such as molecular bulk, or flexibility, to improve the binding constant or to alter the binding kinetics in some desired way. ED has potentially great value for this purpose because it allows the ligand engineer to focus on individual components of the dynamic behavior of the ligand.

Further, it would be very useful to be able to partition the total binding energy among fragments of the ligand, for example, residue by residue in a peptide ligand. Unfortunately, it is impossible to partition binding entropy, and hence the free energy, in this way, as has been pointed out by Mark and van Gunsteren (1994). But the internal energy

can be partitioned, at least formally and for pairwise additive forces, between fragments, and we proceed to do this so that at least one component of the free energy can be studied in this way. In a future extension of this work, we shall study cases where ligand mutants have been made and their binding energies determined experimentally, while we propose to calculate the corresponding changes in the partitioned internal energy. Apart from computing trajectories for the various SH3-peptide ligand complexes, we analyze these trajectories using an essential dynamics algorithm (Ichiye and Karplus, 1991; Amadei et al. 1993) implemented within the MACSIMUS package.

## METHODS

### Choice of starting coordinates

There exist a number of published structures for SH3 domains, both in the unliganded form and as complexes with various peptides. We have chosen to study one such complex determined by x-ray crystallography, the Abl tyrosine kinase SH3 domain, 1abo, which was determined to 2.0-Å resolution by x-ray crystallography (Musacchio et al., 1994). The 1abo file was edited to generate a starting structure file containing a single SH3 (chain A) and peptide ligand (chain C).

Two versions of the ligand were considered. The version in the Protein Data Bank (PDB) file is a zwitterion with the N-end positively charged and the C-end negatively. This version is given the letter Z in this work. The neutral version (referred to as N) the N-end is acetylated ( $\text{CH}_3\text{-C:O-}$ ) and the C-end is substituted by methylamine ( $\text{-NH-CH}_3$ ). These conversions are performed as one of the options in the PDB  $\rightarrow$  MACSIMUS converting utility, which includes energy optimizations of the newly added atoms.

### Force field

The forces between protein molecules are assumed to consist of the following components:

1. Short-ranged nonbonded forces between individual atoms modeled by the Lennard-Jones potential with smooth cutoff, additive diameters, and the Slater-Kirkwood formula for the energy parameters (Brooks et al., 1983).
2. Coulomb nonbonded forces between effective atomic partial charges  $q_i$  and  $q_j$ .
3. Bond potential between a pair of bonded atoms approximated by a harmonic oscillator or a constrained bond.
4. Bond angle bending forces between triplets of particles taking into account nonrigidity of bond angles.
5. Torsional forces between quadruples of bonded atoms to take account of
  - a. the ability of two parts of a molecule connected by a bond to rotate around this bond (proper torsion or a dihedral potential), and
  - b. the need to preserve planarity of  $\text{sp}^2$  hybridized group or chirality of carbon atoms.

Apart from the Coulomb forces, the mathematical forms of these contributions are invariant across various academic and commercial implementations, but different implementations may use different numerical values of the atom-atom specific parameters appearing in those formulas that have usually been found semi-empirically elsewhere in the literature. For the Coulomb forces, the numerical values of the partial charges may likewise vary from one implementation to another. In addition, there are a number of approaches to the problem of how to deal with the long-ranged nature of the Coulomb forces. These include ignoring the problem by just cutting off the potential; introducing an effective dielectric constant into

the Coulomb potential; and using truly periodic potential and replacing the bare Coulomb potential by the periodic Ewald (Allen and Tildesley 1987; de Leeuw et al., 1980) or fast multipole approximation (Greengard 1988)

To elucidate the role of the electrostatic approximations, we implemented both the cutoff version (referred to as C) and the Ewald summation (E). The particular form of the cutoff potential for MD includes smoothing and shifting,

$$U_{\text{elst}}(r) = q_1 q_2 \begin{cases} 1/r - \Delta U & \text{for } r < 0.9r_c \\ (r - r_c)^3(A + Br) & \text{for } 0.9r_c < r < r_c \\ 0 & \text{for } r > r_c, \end{cases} \quad (1)$$

where  $\Delta U$ ,  $A$ , and  $B$  were set according to the conditions of continuous energy  $U_{\text{elst}}(r)$  and forces  $-dU_{\text{elst}}(r)/dr$ .

The Lennard-Jones potential has a smooth cutoff as well (but not shifted) with standard corrections added. Because this potential decays rapidly, there may hardly be any doubt regarding the validity of this approximation in bulk liquid simulations. For details see Kolafa (1999).

Along with the Coulomb force approximations, there are several alternative strategies for dealing with the solvent (water): treating the solvent as vacuum or dielectric continuum; Brownian dynamics, where the dynamic effects of the solvent are modeled by random forces; and full molecular model of the solvent.

It is usual to speak of the set of numerical parameters appearing in the interatomic forces formulae, often with the method used to handle the Coulomb contributions, solvent, etc., used in package X as the “X force field.” The values of partial charges and to some extent also the values of the parameters in the non-Coulomb forces, however, reflect the version of the Coulomb and solvent methodology used. Any comparison of results obtained using different methodology (e.g., cutoff electrostatics vs. Ewald summation) must be made with caution.

To make comparison with other works simple, we used the same values of bonded and nonbonded forces as in the CHARMM 19 force field<sup>2</sup>, (hyperlink <http://www.pharmacy.ab.umd.edu/~alex/>, <http://www.pharmacy.ab.umd.edu/~alex/>) with minor modifications. Aliphatic hydrogens were considered implicitly via the extended (united) atom model for groups CH<sub>n</sub>.

## Simulation conditions

Simulations were performed on a single SH3–ligand complex in a cubic periodic box of 3000 TIP3 waters at density 1 g cm<sup>−3</sup>, which guarantees a large enough box to prevent the protein from interacting with its periodic images. The start-up algorithm contained five steps: 1) energy minimization with all known atom positions fixed and optimized positions of hydrogens (not present in the original PDB files) and atoms in the acetyl and methylamine groups added to cap the termini, 2) random shooting of water molecules into the box with excluded water–protein overlaps, 3) Monte Carlo simulation of the water subsystem (i.e., with the configuration of the protein fixed) to remove water–water overlaps, 4) molecular dynamics equilibration of the water subsystem at temperature  $T = 300$  K, and (5) assigning the atoms of the protein random velocities according to  $T = 300$  K.

Both simulations (N and Z) were carried out twice, once with the use of the Ewald-sum method of handling electrostatics (E) and once using cutoff method (C). Optimal parameters for the Ewald summation were determined according to the error formulae by Kolafa and Perram (1992) with maximum error of force in real space of 0.024 cal/mol/Å and in the reciprocal-space 0.24 cal/mol/Å. This defines two conditions for three parameters, the remaining condition was performance optimization. The final values of parameters are real-space cutoff  $r_c = 18.0$  Å, reciprocal-space cutoff  $K = 8.6$ , and the separation parameter  $\alpha = 0.19/\text{Å}$ . The dielectric constant of continuum surrounding the periodic sample at infinite distance was 80. The value of the cutoff in Eqn. 1 was  $r_c = 12$  Å.

All four simulation runs were carried out for at least 1 ns with 2 fs timesteps. Bond angles containing hydrogens and all bond lengths were constrained using the SHAKE algorithm with the Verlet integrator. The MD simulations were performed using the MACSIMUS package (Kolafa 1999). This shareware package contains code suitable for MD of large and small molecules in various boundary conditions as well as numerous tools for data analysis and visualization (X11 and DOS). It supports spatial decomposition of the simulation box using a variant of the linked-cell list method (Allen and Tildesley, 1987), parallel execution on multiple processors based on this decomposition, and efficient implementation of the Ewald summation. The most demanding calculations using the Ewald summation were performed at the Mærsk McKinney Møller Institute for Production Technology on SGI Onyx parallel supercomputer with 24 R4400 processors, whereas the version with cutoff electrostatics was performed on a standard workstation.

## Principal radii of gyration

Let us consider the inertia tensor  $\mathbf{I}$  and its diagonalization,

$$\mathbf{I} = \sum_{i=1}^N m_i (\mathbf{r}_i - \mathbf{r}_{\text{CM},i})^2 = \mathbf{U} \mathbf{I}_{\text{diag},i} \mathbf{U}^{-1},$$

where  $\mathbf{r}_{\text{CM}}$  is the center of mass and  $\mathbf{U}$  is an orthonormal matrix describing the orientation of the principle axes. The eigenvalues  $\mathbf{I}_{\text{diag},i}$  may be used to define the principle radii of gyration,  $\mathbf{R}_i = (\mathbf{I}_{\text{diag},i}/M)^{1/2}$ , where  $M$  is the total mass. The (total) radius of gyration is then given by

$$\mathbf{R}_g^2 = \mathbf{R}_1^2 + \mathbf{R}_2^2 + \mathbf{R}_3^2.$$

The physical meaning of  $\mathbf{R}_i$  is that a “cloud” of mass  $M$  and Gaussian distribution of density with standard deviations  $\mathbf{R}_i$  in the directions of the principle axes has the same inertia tensor. The three values of  $\mathbf{R}_i$  thus give the first estimate of the shape of the molecule.

## Essential dynamics

The essential dynamics (Amadei et al., 1993) method represents a powerful tool to separate the essential structural changes during the time development from thermal noise. It is based on the analysis of the positive definite, symmetric covariance matrix,

$$\mathbf{C} = \langle (\mathbf{q} - \langle \mathbf{q} \rangle)(\mathbf{q} - \langle \mathbf{q} \rangle)^T \rangle, \quad (2)$$

where  $\mathbf{q}$  denotes the  $3n_{\text{ess}}$ -dimensional vector of  $(x, y, z)$  coordinates of a certain subset of  $3n_{\text{ess}}$  atoms chosen from the studied molecule and  $\langle \cdot \rangle$  is the ensemble average approximated by the average over the MD trajectory.

In principle the essential dynamics algorithm should include all atoms of the studied molecule. For simplicity, a reduced set of atoms is often used in the calculations. We have tried four different subsets of atoms: **C**,  $\alpha$  carbons, for the capped (N) ligand also methyl carbons in acetyl and methylamine; **B**, backbone atoms; **H**, all heavy atoms; and **A**, all atoms (with the exception of aliphatic hydrogens). The coordinates,  $\mathbf{q}$ , entering Eq. 2 are raw coordinates from the simulation only for the first configuration (frame). All subsequent frames were first translated and rotated as to superpose the protein onto the protein in the first frame. This was accomplished by minimizing the mean square distance of both configurations,

$$d(\mathbf{q}_1, \mathbf{q}_i) = \min_{R,T} |\mathbf{q}_1 - TR\mathbf{q}_i|/n_{\text{ess}}, \quad i > 1, \quad (3)$$

where  $R$  and  $T$  represent the operators of rotation and translation, respectively. The minimization thus takes place over six variables (three in the vector of translation  $T$  and three angles in  $R$ ) and was implemented by a Monte Carlo method.

**TABLE 1** Four simulation runs

Symbol	Coulomb	Ligand Ends	t (ns)	d(0, t) (Å)	$\lambda_{1(B)}$	$\lambda_{2(B)}$	$\lambda_{3(B)}$
EZ	Ewald	charged	1.197	2.97	0.407	0.130	0.122
EN	Ewald	capped	1.201	3.60	0.690	0.261	0.194
CZ	cutoff	charged	1.097	2.78	0.202	0.099	0.038
CN	cutoff	capped	1.256	3.05	0.318	0.173	0.106

t, is the run length; d, the mean square distance, Eq. 1; and  $\lambda_{i(B)}$ , are the essential eigenvalues for the backbone atom set.

The eigenvalues  $\lambda_i$  and eigenvectors  $\mathbf{Q}_i$  of the covariance matrix are given by

$$\mathbf{C} \cdot \mathbf{Q}_i = n_{\text{ess}} \lambda_i \mathbf{Q}_i, \quad (4)$$

and are found numerically using the threshold Jacobi diagonalization method (Ralston 1965). The eigenvectors are orthogonal in the  $3n_{\text{ess}}$ -dimensional space. Normalization factor  $n_{\text{ess}}$  makes comparison of results with different atom sets easier.

The eigenvectors  $\mathbf{Q}_i$  corresponding to the highest (or several highest) eigenvalues give the direction of the essential motion relative to the center (average) position  $\langle \mathbf{q} \rangle$ . The quantity of physical interest is the time evolution of the projection of  $\mathbf{q} - \langle \mathbf{q} \rangle$  to  $\mathbf{Q}_i$ , i.e., the amplitude of the  $i$ th essential

motion as the function of time,

$$A_i = \frac{\mathbf{Q}_i \cdot (\mathbf{q} - \langle \mathbf{q} \rangle)}{|\mathbf{Q}_i| n_{\text{ess}}^{1/2}}. \quad (5)$$

The normalization factor  $n_{\text{ess}}^{1/2}$  comes from averaging the (squared) amplitudes over all atoms.

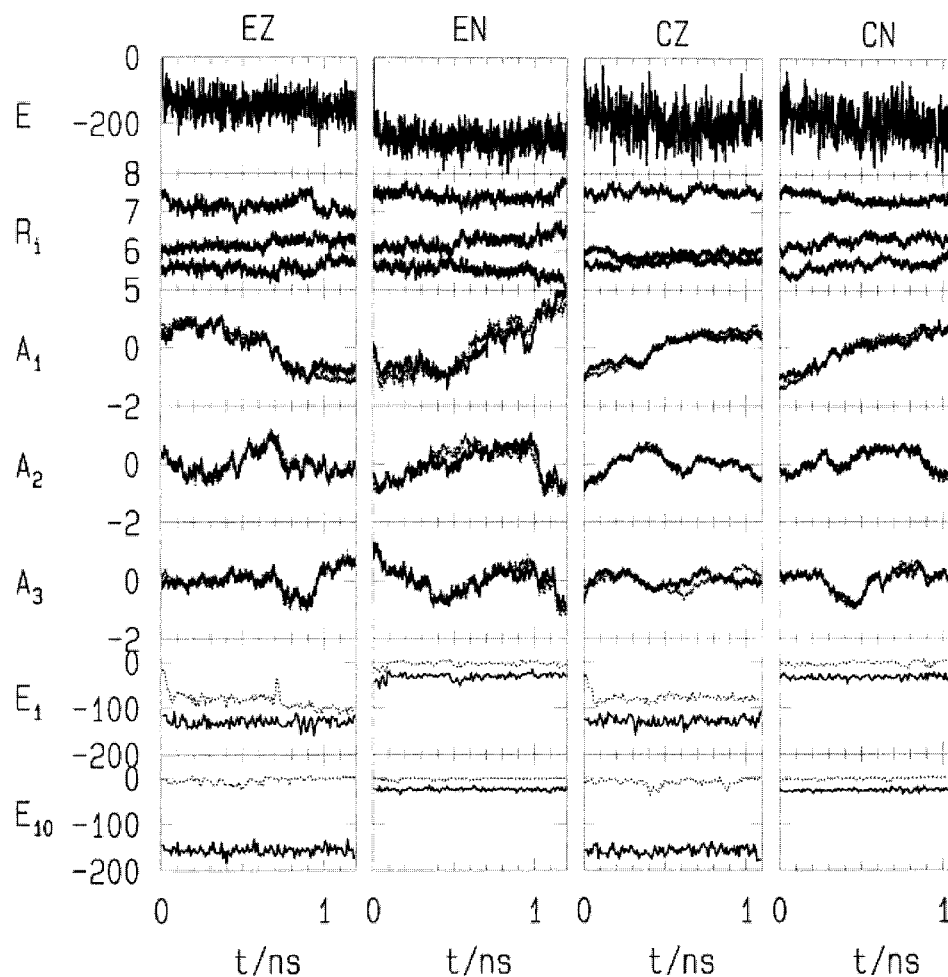
It is useful to be able to compare two essential eigenvectors, either corresponding to different trajectories (simulations runs), or to the same trajectory with different atom sets. The simplest criterion is to compare 3D vectors of essential motions of one chosen atom. The cosine of the angle of these vectors is a measure of similarity—values close to 1 or -1 correspond to parallel motions (note that the sign cannot be distinguished). However, it makes sense to compare signs of cosines calculated for different atoms). Before doing this, it is necessary to match the averaged configurations  $\langle \mathbf{q} \rangle$  for both trajectories, which is done in the same way as in Eq. 3.

Alternatively, it is possible to measure the cosine of the angle in the whole  $3n_{\text{ess}}$  space,

$$c_i = \cos \angle(\mathbf{Q}_i^{(1)}, \mathbf{Q}_i^{(2)}) = \frac{\mathbf{Q}_i^{(1)} \cdot \mathbf{Q}_i^{(2)}}{|\mathbf{Q}_i^{(1)}| |\mathbf{Q}_i^{(2)}|}. \quad (6)$$

If the sets of atoms are different for both vectors, it is necessary first to exclude superfluous atoms to obtain vectors of the same length. This

**FIGURE 1** Convergence profiles of selected quantities for systems EZ, EN, CZ, and CN.  $E$ , total energy in kcal/mol;  $R_g$ , radius of gyration in Å;  $A_i$ , amplitude of  $i$ th essential motion in Å (solid, backbone atoms; dotted, heavy atoms);  $E_i$ , partial residue energy (solid, bare; dotted, solvated). Sampling rates for  $E$  and  $R_g$  are 0.05 ps and averages over 1 ps are shown. Sampling rates of  $A_i$  are 1 ps and of  $E_i$  are 10 ps.





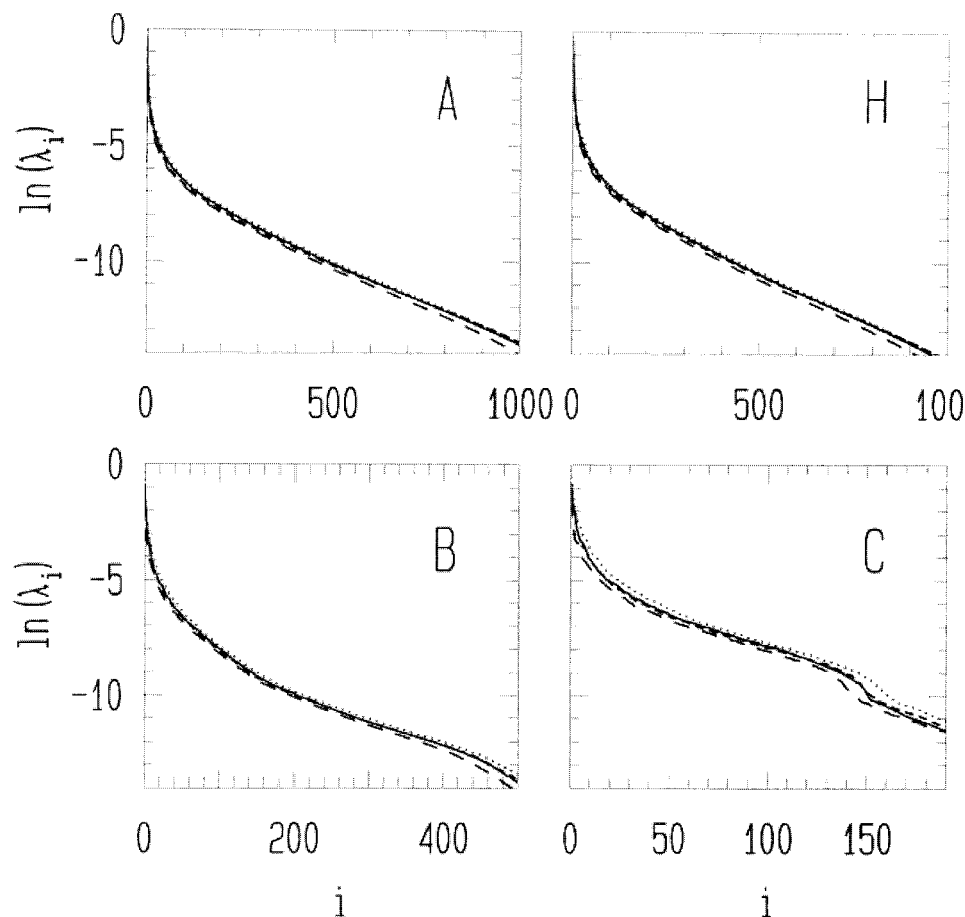


FIGURE 2 Logarithms of essential eigenvalues sorted from the highest to the lowest. Labels A, H, B, and C denote atom sets. *Solid line*, system EZ; *dotted line*, EN; *long-dashed line*, CZ; *short-dashed line*, CN.

criterion is much stronger than the first one because there is much lower probability that two randomly chosen vectors in a many-dimensional space are parallel. This formula can be interpreted also as the correlation coefficient of the set of all coordinates of the motion in the direction of  $\mathbf{Q}_i^{(1)}$  correlated with the corresponding coordinates of  $\mathbf{Q}_i^{(2)}$ . Correlation coefficients between different simulation runs may be defined in the same way. Note that we cannot distinguish  $\mathbf{Q}_i$  from  $-\mathbf{Q}_i$  and thus the sign of  $c_i$  is irrelevant.

### Energy partitioning

To cast light on the ligand–receptor binding mechanisms, we attempted to partition the binding internal energy of individual ligand residues to the receptor in three ways.

1. The “bare” energy is a sum of all pair (Lennard–Jones plus charge–charge) energy contributions between the selected ligand residue and the receptor; residue–residue interactions and receptor–receptor interactions are not included, neither are interactions including water.

2. The “solvated” energy is the bare energy plus the sum of all pair energies between the ligand residue and water.

3. To determine the role of solvation, we also ran short (50 ps) trajectories of pure ligands in water (1000 molecules) and calculated the solvation energies of the ligand residues and water. Both the zwitterionic and capped forms were considered severally with the Ewald summation and cutoff electrostatics.

## RESULTS AND DISCUSSION

### Analysis of trajectories

The ligand bound to the SH3 protein labo is the very proline-rich decapeptide APTMPPPLPP. To judge from the atom types present in the PDB file, this peptide was unprotected at the C- and N-termini, which means that the peptide is a zwitterion. The significance of choosing this particular construct may be questioned if the intention was to mimic the binding of an SH3 domain to a proline-rich stretch of polypeptide chain in, e.g., an SH2 domain in a downstream signaling complex. Such a polypeptide would be “protected” in the sense that it will be flanked by peptide residues of the rest of the main chain of the protein. As such, the peptide groups joining the decapeptide to the rest of the protein would be uncharged, whereas in the unprotected decapeptide (Musacchio et al., 1994), the N-terminal  $\alpha$ -amino group will be positively charged and the C-terminal carboxyl group will be correspondingly negatively charged. We have carried out the simulation of the peptide–protein complex in this form, even though these should presumably be capped. For comparison, we also ran the capped forms.

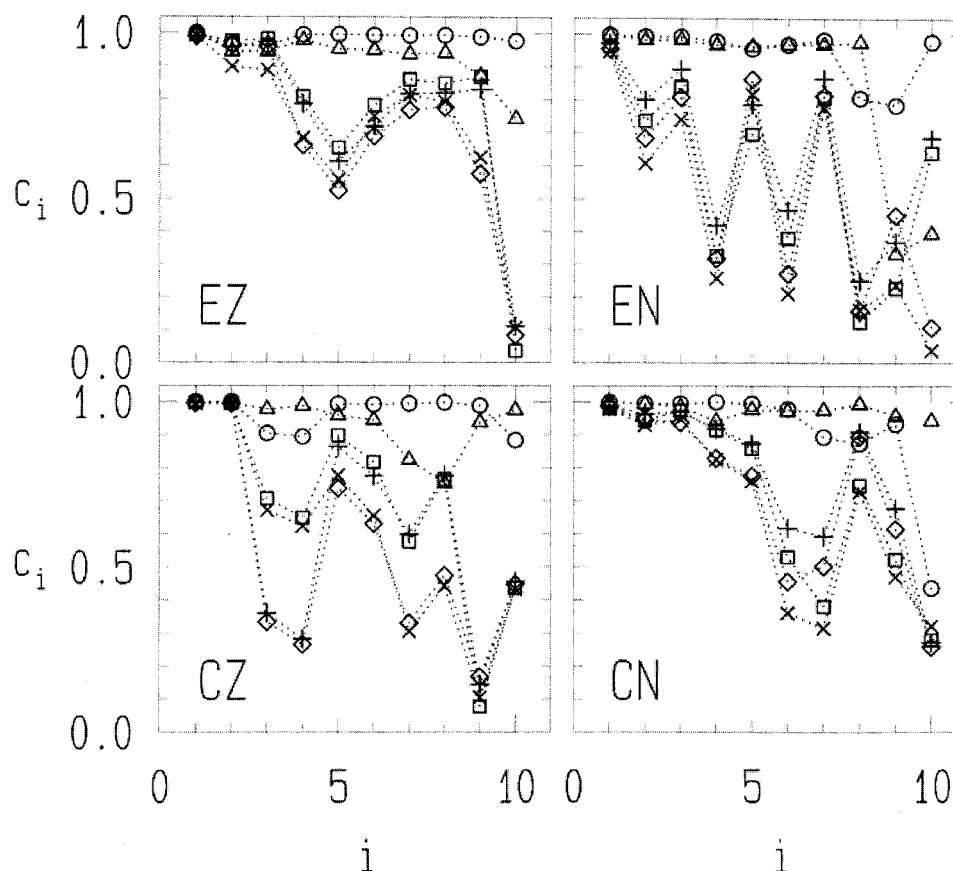


FIGURE 3 Absolute values of the cosines of angles, Eq. 6, between essential motions calculated using different atom sets.  $\circ$ , pair C-B;  $\square$ , C-H;  $\times$ , C-A;  $+$ , B-H;  $\diamond$ , B-A;  $\triangle$ , H-A.

From the runs carried out here, a first impression is that the charged ends of the peptide ligand may have some effect on the ability of the peptide to fully bind to the protein. In both the E and C runs, the C-terminus seems to want to detach itself from the protein, and water molecules are seen to intervene between the partly detached end of the peptide and the protein. This trend was more clearly marked in the capped form. The N-terminus moves during the simulations from the vicinity of TRP110 in the protein to GLU86 (EZ run) or ASN78 (CZ run), whereas in the capped form, it releases. In the CZ run, the unprotected charged ends of the peptide get close together, which inevitably affects the

whole configuration. The largest rearrangement of the peptide occurs for the EN run. This is also reflected by the mean-squared distance of the starting and ending configuration, see Table 1.

Time development of the protein structure is characterized by convergence profiles of selected quantities, which are collected, for the four trajectories considered, in Fig. 1, frames 1–4. Shown variables are total energy  $E$ ; principle radii of gyration  $R_i$ ; amplitudes of the first three essential motions,  $A_1$ ,  $A_2$ , and  $A_3$ ; and partial energies of the first and last residue  $E_1$  and  $E_{10}$ .

## Essential Dynamics

The first question, though very technical, is how the choice of the atom set in the essential analysis affects the results. The simplest information is seen in Fig. 2, which shows the spectra of essential eigenvalues for four sets of atoms. For the lower indices, up to 5–10, the curves for  $\lambda_i$  versus  $i$  for all four sets are almost parallel. This suggests that the essential motions affect the whole backbone and are sufficiently described by the sparsest atom set C. For higher indices, curves C and B gradually begin to deviate from curves A and H, exemplifying the role of motions of sidechains in the latter case. Interesting and not understood

TABLE 2 Comparison of the most significant essential motion between simulation runs

Pair of Systems	ALA1	PRO6	PRO10	cos (angle)
CN-CZ	-0.928	-0.569	-0.015	0.4034
CN-EN	0.863	-0.414	0.176	-0.0785
CN-EZ	0.021	-0.706	-0.513	-0.3305
CZ-EN	-0.787	-0.344	-0.984	-0.3826
CZ-EZ	-0.322	0.689	-0.672	-0.1150
EN-EZ	0.271	-0.305	0.539	0.1431

Columns ALA1, PRO6, and PRO10 show cosines of angles between the motions of  $C_\alpha$  atoms in the respective residues and  $\cos(\text{angle})$ , Eq. 6, is the angle for the B set of atoms (extra atoms in the N systems were removed).

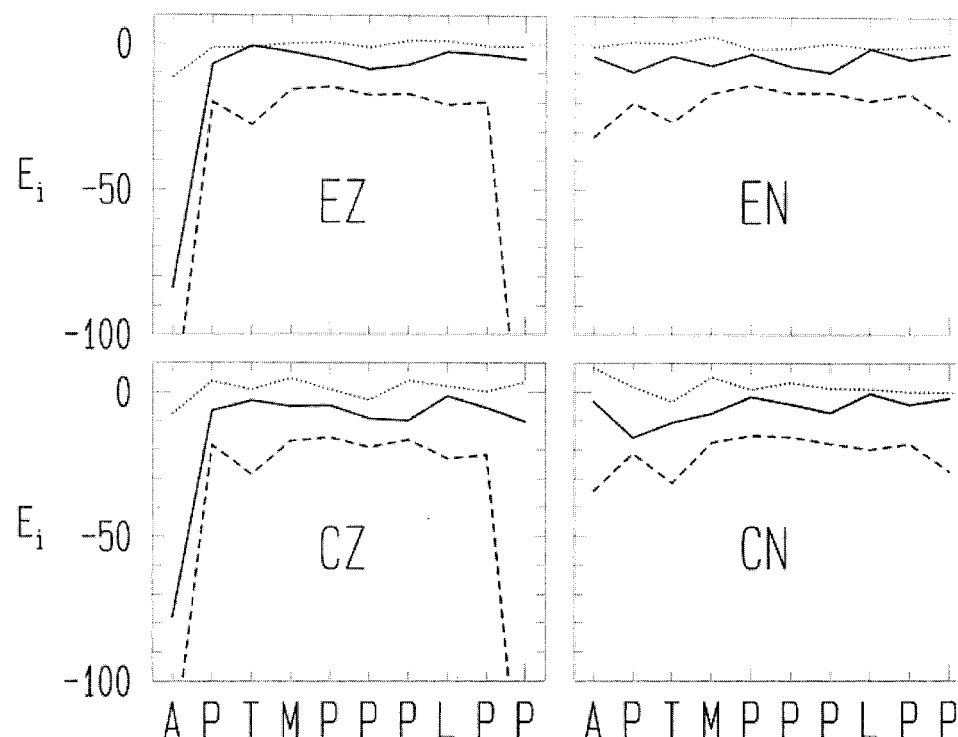


FIGURE 4 Energy partitioning of ligand-receptor interactions. The  $x$  axis shows the residue. *Solid lines*, bare residue-receptor energies; *dashed lines*, solvated energies; and *dotted lines*, solvated energies minus solvated energies of the respective free ligand in water.

is the decrease of  $\lambda_i(C)$  at  $i$  about 130–140. Note that the C-spectrum ends at  $i_{\text{end}} = 3(n_{\text{ess}} - 2) = 198$  for Z systems and  $i_{\text{end}} = 204$  for N; it holds  $\lambda_i = 0$  for  $i_{\text{end}} < i \leq 3n_{\text{ess}}$  because these last six eigenvalues correspond to rotations and translations that have been filtered out. To elucidate this question, we plotted in Fig. 3 cosines of angles between eigenvectors for different atom sets, Eq. 6; all vectors entering this equation are reduced to positions of the  $C_{\alpha}$  atoms.

It is seen that the first several (2–4) essential motions are always highly correlated, in other words, the C set is sufficient. The correlations C versus B (only backbone) and A versus H (side chains) are higher than for any combination of an atom set containing only backbone (C and B) versus an atom set with side chains (A and H). For system CZ, the third most significant motion takes place in side chains.

Along with Fig. 2, numerical values of three most significant eigenvalues are also shown in Table 1. The highest value is reached for EN system. This can be interpreted such that the motions in this system occur mainly along one direction, or that one (linear) mode of motion is by far most significant. Note that, for this system also, the averaged mean distance  $d$  reaches maximum, i.e., this system is subject to the biggest structural changes.

The time development of eigenvalues and of the partial radii of gyration reveal one important feature of the simulations, namely, that the complex undergoes a gradual change of conformation rather than fluctuations from a well-defined equilibrium value. In addition, the results of all

four runs seem to be uncorrelated, see Table 2 and Fig. 1. This may reflect 1) inaccurate force field, 2) inappropriate simulation methodology, 3) inappropriate ligand model, or 4) the system has not reached equilibrium. As regards 1, there is not too much space left for future improvements in the current state-of-the-art of complex molecule simulation technology. One might include explicit aliphatic hydrogens, try a different force field, or include a better water model. As regards 2, we have tested the Ewald summation. This is supposed to be more accurate than cutoff electrostatics, provided that the values of partial charges are adjusted accordingly, which is questionable. As a consequence of omitting some parts of the Coulomb interaction, the cutoff method destabilizes the structure as can be deduced from the larger fluctuations of the energy curves in Fig. 1. As for 3, the short decapeptide is an approximation of the SH2 ligand loop, which, in complexes involving the intact protein, may be much more stable. The capped version of the model peptide (presumably more appropriate as a model of the real complex) tends to detach more easily than the zwitterionic form. This may reflect the fact that SH2–SH3 complexes need to dissociate at some stage, in addition to binding. The charges at the end of the zwitterionic form introduce attractive interactions not present in the natural SH2–SH3 complex. As regards 4, which is suggested by the convergence profiles in Fig. 1, correlation times for protein motions are of the order of milliseconds. Therefore, the only way to assess the accuracy of short trajectories is to compare several independent runs.

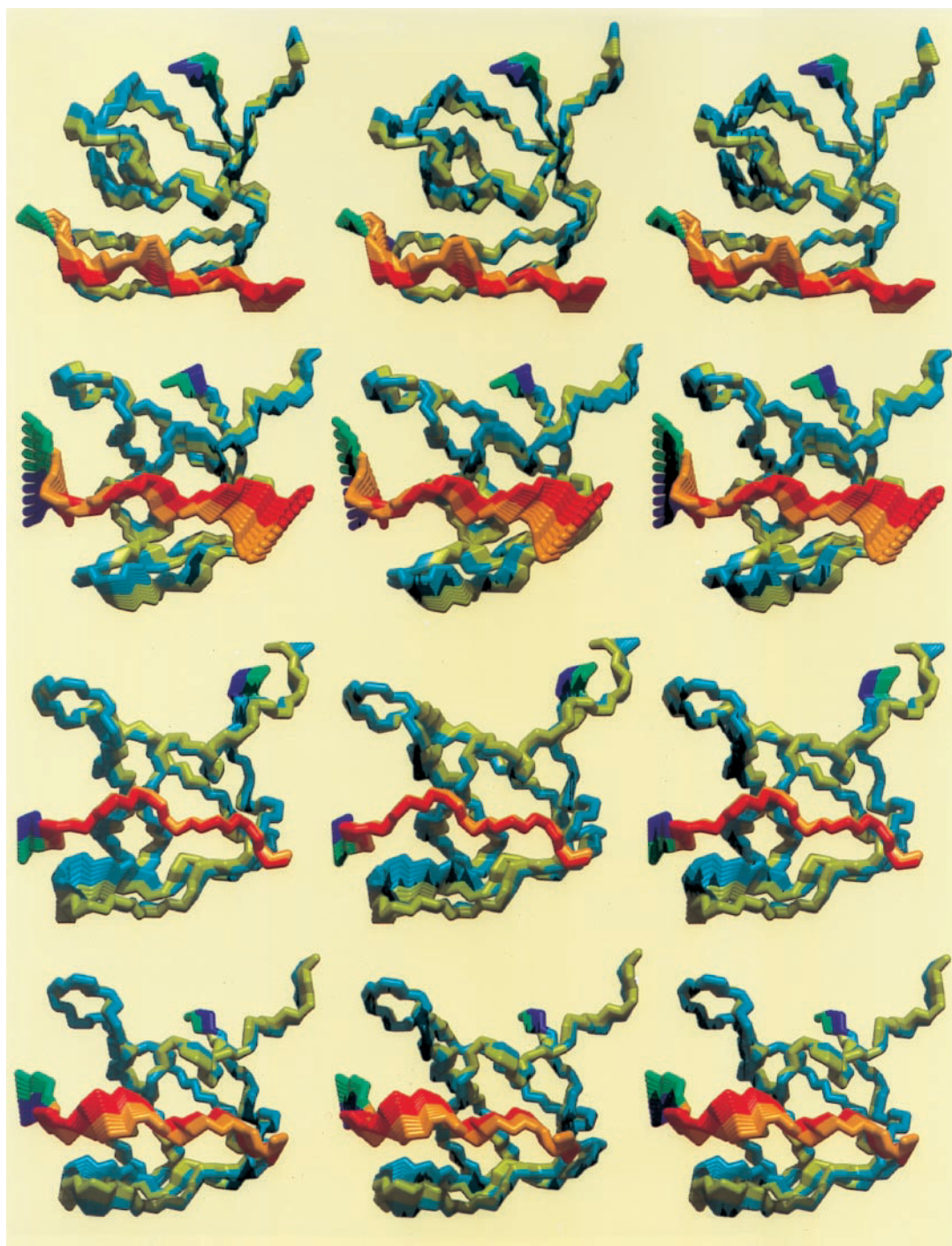


FIGURE 5 Stereo picture of the backbone essential motion corresponding to the highest eigenvalue for systems from the top EZ, EN, CZ, and CN. Color changes from yellow to orange (the ligand) and from cyan to green correspond to the same direction of motion. The N-termini are blue and green. The amplitude of the motion is given by the condition of the same mean square displacement as the real motion in the direction of the essential eigenvector. The left and right figures are for the left eye and the central figure for the right eye. Watch the left pair with parallel eyes (the cross-point is behind the paper) or the right pair with eyes crossed at a point between the paper and eyes.

In terms of accuracy, i.e., ability to reproduce experimental structural data, the x-ray crystal structure may differ from the structure (or rather set of dynamic structures) found in solution. First, crystal structures are a time average

and second, by their very nature, the molecular structures are subject to perturbations arising from crystal packing forces. Another factor that must be born in mind when comparing crystal structures and those obtained by molec-



ular simulations is that the crystal structures are refined using energy minimization software, which may have a different force field than the one used in the simulations.

### Partitioning of the energetic contributions from each residue

The results for the four cases EZ, CZ, EN, and CN are shown in Fig. 4. The two zwitterionic cases behave very similarly, with much lower energy contributions from the terminal residues. This is most apparent when water–ligand interactions are included. When the water contribution is absent, the C-terminal end has an energy contribution more in line with the other residues. This may reflect the observation above that, in the trajectories, the C-terminus appears to want to detach itself and admit water. For the residues along the chain, only threonine, in the third position, shows any behavior that deviates markedly from the other residues, and this is only apparent when water is included. EN and CN behave similarly to each other, but, compared with the zwitterionic pair, show much more marked differences as one proceeds along the chain. When water is included, the termini show a low energy contribution and the threonine behaves as before with the leucine in position 8 also deviating somewhat.

The dotted lines in Fig. 5 are obtained by subtracting solvation energies of the respective residues of a free ligand in water from the solvation energies of a bound ligand. Even though the solvation of a free ligand may be affected by random conformational changes (there results were obtained by independent MD runs), this difference represents “clean” binding energies of individual ligands in water environment. If we consider the size of the residues and a special role of the ends, the results support the contention that the prolines contribute most to the binding.

We have not yet investigated what happens if residues in the peptide are mutated. This kind of simulated experiment will be carried out in other peptide–protein systems where there is experimental data (binding energies) for the binding of different mutants. Some data of this kind has recently been published by Nguyen et al. (1998), who showed that not only proline, but other N-substituted residues, can bind very efficiently to SH3. It would be of interest to include these mutants in future simulation studies.

## CONCLUSIONS

### Analysis of trajectories

The ends of the peptide tend to bind more loosely in the capped form than in the zwitterionic form, due to the lack of the electrostatic interactions operating in the latter case. Inasmuch as we surmise that the capped form is more relevant to the real situation where SH3 recognizes SH2, this suggests that the end residues in the peptide fragment

are not important for binding, whereas the middle residues, including proline, are responsible for most of the binding energy.

### Essential dynamics

It is possible to observe significant motions in the bound ligand and to distinguish between motions of the peptide backbone and of the side chains. This could be useful in designing ligands, which fit optimally to the binding protein. The most favorable model for observing these motions is probably EN, i.e., the capped peptide studied using Ewald summation. From the practical point of view, it is important that the first essential motions are sufficiently well described by the atom set reduced to the  $C_\alpha$  atoms only.

### Partitioning of the energetic contributions from each residue

We show that it is possible to determine the different contributions of each residue in a peptide to the enthalpy of binding. Proline is probably the major contributor to binding enthalpy, in keeping with the observed propensity for this family of proteins to bind proline-rich peptides.

The authors are grateful to the Danish Science Research Councils' programs in information technology (PIFT). The simulations were performed on the 24 processor SGI Onyx supercomputer, which was purchased with the financial support of the Danish Research Councils, DOU, Novo Nordisk A/S, Odense Steel Shipyard, and Odense University Faculty of Natural Sciences. The development of MACSIMUS was partly supported also by the LBL-DOE program on advanced batteries, and the ARO.

## REFERENCES

- Allen, M. P., and D. J. Tildesley. 1987. *The Computer Simulation of Liquids*. Clarendon Press, Oxford, U.K.
- Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.
- Bækdal, L., W. Joosen, T. Larsen, J. Kolafa, J. H. Ovesen, J. W. Perram, H. G. Petersen, R. Bywater, and M. Ratner. 1996. The object oriented development of a parallel application in protein dynamics: why we need software tools for HPCN applications. *Comput. Phys. Com.* 97:124–135.
- Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer based archival file for macromolecular structure. *J. Mol. Biol.* 112:535–542.
- Barfod, E. T., Y. Zheng, W. J. Kuang, M. J. Hart, T. Evans, R. A. Cerione, and A. Ashkenazi. 1993. Cloning and expression of a human CDC42 GTPase-activating protein reveals a functional SH3-binding domain. *J. Biol. Chem.* 268:26059–26062.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
- Cicchetti, P., B. J. Mayer, G. Thiel, and D. Baltimore. 1992. Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science*. 257:803–806.

- de Leeuw, S. W., J. W. Perram, and E. R. Smith. 1980. Simulations of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. Lond. A*. 373:27–56.
- Feng, S., J. K. Chen, H. Yu, J. A. Simon, and S. L. Schreiber. 1994. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science*. 266:1241–1247.
- Fumagalli, S., N. F. Totty, J. J. Hsuan, and S. A. Courtneidge. 1994. A target for Src in mitosis. *Nature*. 368:871–874.
- Gout, I., R. Dhand, I. D. Hiles, M. J. Fry, G. Panayotou, P. Das, O. Truong, N. F. Totty, J. Hsuan, G. W. Booker, I. D. Campbell, and M. D. Waterfield. 1993. The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell*. 75:25–26.
- Greengard, L. 1988. *The Rapid Evaluation of Potential Fields in Particle Systems*. MIT Press, Cambridge, MA.
- Ichiye, T., and M. Karplus. 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal modes simulations. *Proteins Struct. Funct. Genet.* 11:205–217.
- Kato, J. Y., T. Takeya, C. Grandori, H. Iba, J. B. Levy, and H. Hanafusa. 1986. Amino acid substitutions sufficient to convert the nontransforming p60c-src protein to a transforming protein. *Mol. Cell Biol.* 6:4155–4160.
- Koch, C. A., D. Anderson, M. F. Moran, C. Ellis, and T. Pawson. 1991. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science*. 252:668–674.
- Kolafa, J. 1999. MACSIMUS (<http://www.icpf.cas.cz/jiri/macsimus>)
- Kolafa, J., and J. W. Perram. 1992. Cutoff errors in the Ewald summation formulae for point charge systems. *Mol. Sim.* 9:351–368.
- Liu, X., L. E. Marengere, C. A. Koch, and T. Pawson. 1993. The v-Src SH3 domain binds phosphatidylinositol 3'-kinase. *Mol. Cell Biol.* 13: 5225–5232.
- Mark, A. E., and W. F. Van Gunsteren. 1994. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. *J. Mol. Biol.* 240:167–176.
- McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of folded proteins. *Nature*. 267:585–590.
- Murzin, A., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540. (<http://scop.mrc-lmb.cam.ac.uk/scop/>)
- Musacchio, A., T. Gibson, V. P. Lehto, and M. Saraste. 1992. SH3—an abundant protein domain in search of a function. *FEBS Lett.* 307:55–61.
- Musacchio, A., M. Saraste, and M. Wilmanns. 1994. High-resolution crystal structures of tyrosine kinase SH3 domains complexed to proline-rich peptides. *Nature Struct. Biol.* 1:546–551.
- Nemeth, S. P., L. G. Fox, M. DeMarco, and J. S. Brugge. 1989. Deletions, within the amino-terminal half of the c-src gene product that alter the functional activity of the protein. *Mol. Cell Biol.* 9:1109–1119.
- Nguyen, J. T., C. W. Turck, F. E. Cohen, R. N. Zuckermann, and W. A. Lim. 1998. Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science*. 282:2088–2092.
- Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. 1997. CATH—a hierarchic classification of protein domain structures. *Structure*. 5:1093–1108. (<http://www.biochem.ucl.ac.uk/bsm/cath/>)
- Pawson, T., and G. D. Gish. 1992. SH2 and SH3 domains: from structure to function. *Cell*. 71:359–362.
- Pawson, T., and J. Schlessinger. 1993. SH2 and SH3 domains. *Curr. Biol.* 3:434–442.
- Peters, G. H., D. M. F. van Aalten, A. Svendsen, and R. P. Bywater. 1997. Essential dynamics of lipase binding sites: the effect of inhibitors of different chain length. *Protein Eng.* 10:149–158.
- Peters, G. H., and R. P. Bywater. 1999. Computational analysis of chain flexibility and fluctuations in *Rhizomucor miehei* lipases. *Protein Eng.* 12:747–754.
- Potts, W. M., A. B. Reynolds, T. J. Lansing, and J. T. Parsons. 1988. Activation of pp60c-src transforming potential by mutations altering the structure of an amino terminal domain containing residues 90–95. *Oncogene Res.* 3:343–355.
- Ralston, A. 1965. *A First Course in Numerical Analysis*. McGraw-Hill.
- Seidel-Dugan, C., B. E. Meyer, S. M. Thomas, and J. S. Brugge. 1992. Effects of SH2 and SH3 deletions on the functional activities of wild-type and transforming variants of c-Src. *Mol. Cell Biol.* 12:1835–1845.
- Taylor, S. J., and D. Shalloway. 1994. An RNA binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature*. 368: 867–871.
- Van Aalten, D. M. F., A. Amadei, R. Bywater, J. B. C. Findlay, H. J. C. Berendsen, C. Sander, and P. F. W. Stouten. 1996. A comparison of structural and dynamic properties of different simulation methods applied to SH3. *Biophys. J.* 70:684–692.
- Van Gunsteren, W. F., and H. J. C. Berendsen. 1987. GROMOS: Groningen molecular simulation computer program package. University of Groningen, Groningen, The Netherlands.
- Weng, Z., J. A. Taylor, C. E. Turner, J. S. Brugge, and C. Seidel-Dugan. 1993. Detection of Src homology 3-binding proteins, including paxillin, in normal and v-Src-transformed Balb/c 3T3 cells. *J. Biol. Chem.* 268: 14956–14963.
- Weng, Z., S. M. Thomas, R. J. Rickles, J. A. Taylor, A. W. Brauer, C. Seidel-Dugan, W. M. Michael, G. Dreyfuss, and J. S. Brugge. 1994. Identification of Src, Fyn, and Lyn SH3-binding proteins: implications for a function of SH3 domains. *Mol. Cell Biol.* 14:4509–4521.